Microbiological Research 238 (2020) 126525

Contents lists available at ScienceDirect

Microbiological Research

journal homepage: www.elsevier.com/locate/micres

New insights into the variability of lactic acid production in *Lachancea thermotolerans* at the phenotypic and genomic level

Veronica Gatto[a,1], Renato L. Binati[a,1], Wilson J.F. Lemos Junior[a], Arianna Basile[b], Laura Treu[b], Otávio G.G. de Almeida[c], Giada Innocente[a], Stefano Campanaro[b], Sandra Torriani[a,*]

[a] Department of Biotechnology, University of Verona, 37134, Verona, Italy
[b] Department of Biology, University of Padua, 35121, Padua, Italy
[c] Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, 14040-900, Ribeirão Preto, Brazil

ARTICLE INFO

Keywords:
Lachancea thermotolerans
L(+)-Lactic acid
Intraspecific variability
Biological acidification
Whole-genome sequencing
Gene expression

ABSTRACT

Non-conventional yeasts are increasingly applied in fermented beverage industry to obtain distinctive products with improved quality. Among these yeasts, *Lachancea thermotolerans* has multiple features of industrial relevance, especially the production of L(+)-lactic acid (LA), useful for the biological acidification of wine and beer. Since few information is available on this peculiar activity, the current study aimed to explore the physiological and genetic variability among *L. thermotolerans* strains. From a strain collection, mostly isolated from wine, a huge phenotypic diversity was acknowledged and allowed the selection of a high (SOL13) and a low (COLC27) LA producer. Comparative whole-genome sequencing of these two selected strains and the type strain CBS 6340^T showed a high similarity in terms of gene content and functional annotation. Notwithstanding, target gene-based analysis revealed variations between high and low producers in the key gene sequences related to LA accumulation. More in-depth investigation of the core promoters and expression analysis of the genes *ldh*, encoding lactate dehydrogenase, indicated the transcriptional regulation may be the principal cause behind phenotypic differences. These findings highlighted the usefulness of whole-genome sequencing coupled with expression analysis. They provided crucial genetic insights for a deeper investigation of the intraspecific variability in LA production pathway.

1. Introduction

Besides *Saccharomyces cerevisiae*, the most studied and exploited yeast species, non-conventional yeasts belonging to different genera and species have recently received increasing attention by the scientific community and the beverage fermentation industry (Capozzi et al., 2015; Osburn et al., 2018; Lemos Junior et al., 2020). Among these yeasts, *Lachancea thermotolerans*, formerly *Kluyveromyces thermotolerans*, has a remarkable technological potential, yet underexplored (Kurtzman, 2003; Hranilovic et al., 2018). Above all, *L. thermotolerans* ability to produce L(+)-lactic acid (LA) during fermentation is a very unusual metabolic activity among yeasts and has gained great biotechnological interest (Witte et al., 1989; Sauer et al., 2010). LA is considered one of the most industrially important organic acids, due to many versatile applications in food, pharmaceutical, textile, chemical industries (Datta and Henry, 2006; Sauer et al., 2008; Chen et al., 2013; Martinez et al., 2013). Two enantiomeric forms are widely distributed in nature: L(+) and D(-), and microorganisms can selectively produce one of two stereoisomers or a racemic mixture by lactic fermentation (Axelsson, 2004; Martinez et al., 2013).

The efficiency of *L. thermotolerans* to produce LA is low for the industrial bulk chemical manufacture of this substance and hence not much effort was put into this yeast as a production organism. However, in some processes the concomitant acidification with alcoholic fermentation is a benefit, notably in the oenological industry context. LA could provide an effective acidification and a higher microbial stability, important for some grape cultivars and wine-producing regions where the acidity is insufficient (Sauer et al., 2010; Jolly et al., 2014; Hranilovic et al., 2017). Several studies have proposed the use of *L. thermotolerans* in multi-starter fermentations with *S. cerevisiae* to improve aroma, flavour and mouthfeel of wine and reduce the ethanol content (Lachance and Kurtzman, 2011; Freel et al., 2014; Benito et al., 2016). Indeed, *L. thermotolerans* contribution could be useful to address some recent concerns of the wine industry regarding climate change,

* Corresponding author.
E-mail address: sandra.torriani@univr.it (S. Torriani).
[1] Both authors equally contributed to this work.

https://doi.org/10.1016/j.micres.2020.126525
Received 15 April 2020; Received in revised form 3 June 2020; Accepted 5 June 2020
Available online 17 June 2020

which is causing an increase of sugar levels (and consequently of ethanol) and loss of acidity of the grapes (Balikci et al., 2016; Benito et al., 2016).

Strains of this species were also proposed as starter in the growing sour beer market, providing organoleptic differentiation, enhancing the aroma complexity and avoiding the addition of bacteria, thereby simplifying and shortening the process (Domizio et al., 2016; Osburn et al., 2018).

Recent studies have underlined significant phenotypic variability among *L. thermotolerans* strains when important technological parameters, such as LA production, are compared. This indicates that a proper selection has to be carried out in order to find the most appropriate cultures for overall product enhancement (Domizio et al., 2016; Benito, 2018).

Despite the great interest in LA production and the information derived by the complete genome sequence of *L. thermotolerans* CBS 6430^T (Génolevures Consortium et al., 2009), which is the only one available for this species to date, the knowledge regarding lactate metabolism, either at phenotypic or genotypic level, is still limited.

In yeasts, the major flux of pyruvate metabolism is to ethanol, a step catalysed by pyruvate decarboxylase (PDC) and alcohol dehydrogenase (ADH). Pyruvate can be reduced to lactate by the activity of lactate dehydrogenase (LDH), providing an alternative route for regenerating NAD^+ and reducing alcoholic fermentation (Hranilovic et al., 2018). LDH belongs to the L-lactate/L-malate dehydrogenase (L-LDH/L-MalDH) super-family. They can be distinguished by the presence of catalytically important amino acid residues for substrate specificity, i.e. glutamine instead of arginine at position 102, and isoleucine instead of serine or proline at position 250. Moreover, the signature sequence GXGXXG close to the N-terminus is thought to be specific for L-LDH (Madern, 2002).

Transport of monocarboxylic acids across the plasma membrane is another essential step of LA production (Makuc et al., 2001). Intracellular accumulation can result in feedback inhibition of LDH, leading to a decrease in LA yield (Pacheco et al., 2012; Turner et al., 2019). The role of the monocarboxylate transporters JEN1 and ADY2, as modulators for LA export by the cell, was well-demonstrated in different studies using engineered *S. cerevisiae* strains (de Kok et al., 2012; Pacheco et al., 2012). It is not yet clear whether JEN1 and ADY2 are the sole transporters carrying out acid efflux or if another LA exporter (not yet identified) is also involved. However, the existence of other exporters cannot be ruled out, as in strains with double deletion of *jen1* and *ady2* the LA efflux was still operational, although reduced (Pacheco et al., 2012; Turner et al., 2019).

Aiming to shed light on this important metabolic feature, a collection of *L. thermotolerans* wine-associated strains was characterized for the ability to produce LA. Afterwards, whole-genome of the highest and lowest LA producer strains was sequenced in order to unravel intraspecific genomic differences. Focusing on the LA production pathway, the most important genes related to synthesis and transport were compared. Finally, to assess potential differences at transcription level, promoter sequences and expression of the genes *ldh* were analysed.

2. Material and methods

2.1. Yeast strains and growth conditions

The *L. thermotolerans* strains used in this study, listed in Table 1, have been previously isolated from several wine grape varieties, characterized and identified by phenotypical and molecular methods (Binati et al., 2019, 2020). The strains DBVPG 6232^T = CBS 6340^T (hereinafter indicated as CBS 6340^T) and Viniflora® CONCERTO™ were also included for comparative purposes. All yeasts were maintained in the Department of Biotechnology culture collection.

The yeasts were reactivated from frozen cultures on Wallerstein Laboratory (WL) nutrient agar (Sigma-Aldrich, Milan, Italy) plates, and then inoculated in YPD broth (10 *g*/L Yeast extract, 20 *g*/L bacteriological Peptone, 20 *g*/L Dextrose) in agitation overnight at 27 °C to reach the early stationary phase. The cultures were centrifuged at 3000 × *g* for 5 min, washed twice, and re-suspended in sterile physiological solution (0.9 % w/v NaCl). The cell suspension of each strain, at about 10^8 cells/mL, was used to inoculate the media for the phenotypic and genotypic assessment of LA production.

Table 1
Code and isolation source of the wine-associated *L. thermotolerans* strains studied. The strains subjected to *de novo* whole-genome sequencing are in bold.

Isolate code	Grape variety	Region of isolation (Province)
COLC11	Pinot Grigio	Veneto (VR)
COLC27	Pinot Grigio	Veneto (VR)
DESP53	Marzemino	Veneto (TV)
FIANO43	Fiano	Puglia (BA)
GLERA15	Glera	Veneto (TV)
LS15	Alicante	Toscana (GR)
LT15	Malvasia Candia	Emilia Romagna (PC)
LT3	Malvasia Candia	Emilia Romagna (PC)
MALV13	Malvasia Candia	Emilia Romagna (PC)
MALV17	Malvasia Candia	Emilia Romagna (PC)
SOL13	Solaris	Trentino Alto Adige (TN)

2.2. L(+)-lactic acid production in mYPD broth and in grape must

LA production by *L. thermotolerans* strains was assayed in 96-well microtiter plates (Greiner Bio-One, Kremsmünster, Austria). Wells were filled in with 198 µL of YPD broth modified with the addition of glucose to reach a final concentration of 220 g/L (mYPD), and inoculated with 2 µL of the cell suspension prepared as described above. All yeasts were inoculated in triplicate. The optical density at 600 nm (OD_{600}) was measured with a Cary 60 UV–vis spectrophotometer (Agilent Technologies, Santa Clara, USA) every 24 h during three days of incubation in order to follow the yeast growth. LA concentration was quantified in the supernatant after 72 h, by using an enzymatic assay (Megazyme, Bray, Ireland), following the manufacturer's protocol.

Two selected *L. thermotolerans* strains (SOL13 and COLC27) were singly inoculated in microvinification trials using biological white grape must (160 g/L of sugar, Folicello, Modena, Italy), pasteurized at 70 °C. Must samples (100 mL) were inoculated with approximately 10^6 cells/mL. Fermentations were carried out in triplicate for each strain at 22 °C under static conditions, and weight loss and OD_{600} were measured daily until the end of fermentation (constant weight). Samples for LA quantification were taken at T1 (half fermentation; 144 h) and T2 (end of fermentation; 288 h). LA concentration was measured enzymatically as indicated above.

2.3. Whole-genome sequencing and comparison

Total genomic DNA from the *L. thermotolerans* strains SOL13 and COLC27 was isolated and purified using the Wizard Genomic DNA Purification kit (Promega, Milano, Italy), according to the manufacturer's instructions. A volume of 50 mL of cultures grown on YPD broth was used to have a sufficient DNA quantity for genome sequencing. DNA yield and purity were estimated by Nanodrop ND1000 UV–vis Spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and Qubit 4 Fluorometer (Thermo Fisher Scientific).

The shotgun sequences were generated using an Illumina NextSeq 500 platform (1-kb paired end libraries) at the CRIBI Biotechnology Center (Padua, Italy). Genome sequences were assembled *de novo* using CLC Genomic Workbench software (version 9.5). The genome quality was assessed with Quast (v5.0.2–1) (Gurevich et al., 2013), and the genome was measured with BUSCO "fungi_odb9" database (Simão et al., 2015). The Whole Genome Shotgun project of the *L.*

thermotolerans strains SOL13 and COLC27 has been deposited at DDBJ/ENA/GenBank under the accession numbers WVSD00000000 and WVSE00000000, respectively.

Coding sequences (CDS) were predicted by GeneMark-ES (v4.57) (Besemer and Borodovsky, 2005). Gene sequences were retrieved from the GTF file through the get_sequence_from_GTF.pl script. Gene functional annotation was performed using the BlastKOALA tool (Kanehisa et al., 2016), to access the non-redundant set of KEGG genes, with the selection of the family Saccharomycetaceae as a taxonomy group and 4893 as taxonomy ID. RPS BLAST was used to compare protein sequences with eukaryotic orthologous groups of proteins (KOG) (Tatusov et al., 2003).

Variants differing between COLC27 and SOL13 were identified by aligning back the reads of one strain on the assembled genome of the other strain. Reads were filtered with Trimmomatic (v0.33) using the following parameters: LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:65 (Bolger et al., 2014). Reads were aligned using Bowtie2 (v2.2.4) (Langmead and Salzberg, 2012). Files in "sam" format were converted to "bam" format using samtools (Li et al., 2009) and finally checked and converted to "vcf" using samtools mpileup (v1.9–12) and bcftools (v1.6). SnpEFF software (v4.3 g) was used for genomic variants annotation and functional effect prediction (Cingolani et al., 2012) using SOL13 as reference genome. Only variants with predicted "high" or "moderate" effect were selected. Function of the genes with variants was obtained by annotating proteins with BlastKOALA software (Kanehisa et al., 2016). Progressive MAUVE (Darling et al., 2004) was used for genome alignment to identify translocations between *L. thermotolerans* CBS 6340^T and the two examined strains targeting the genes *ldh*.

Phylogenomic analysis were performed with all the available genomes of *Lachancea* species in the NCBI database. The annotation was carried out using the Funannotate pipeline choosing the tantan method to identify repetitive regions. The *ab initio* predictions were based on a set of non-redundant proteins from Uniprot KB database. The genomes of *L. thermotolerans* COLC27 and SOL13 were annotated combining GeneMark-ES and Augustus consensus predictions. Gene function was inferred matching Uniprot KB, CAZy, Merops and pre-computed BUSCO dikarya models. The strains orthologs were identified by the get_homologs-est pipeline (Contreras-Moreira and Vinuesa, 2013) with a threshold of 80 % of similarity, aligned using Clustal Omega (Sievers et al., 2011), trimmed with TrimAL and concatenated. The concatenated FASTA file was processed with IQ-Tree software version 1.6.10 (Nguyen et al., 2015) with 1000 Ultra-fast bootstraps (Minh et al., 2013), rendering a Maximum-Likelihood (ML) phylogenetic tree determined with ModelFinder (Kalyaanamoorthy et al., 2017) (Best-fit model TIM2 + F + I + G4 according to the BIC - Bayesian Information Criterium), which was visualized with iTOL v5 software (Letunic and Bork, 2019). The *Schizosaccharomyces pombe* 972 h was used as outgroup species.

The alignment and clustering of amino acid sequences of LDH and monocarboxylate transporters (JEN1 and ADY2) from *L. thermotolerans* strains, and of the promoter regions of the genes *ldh*, were conducted using Clustal Omega (Sievers et al., 2011) and MEGA6 software (Molecular Evolutionary Genetics Analysis) (Tamura et al., 2013). Conserved domains in the amino acid sequences were detected using BlastP protein database search program version 2.10.0 (Altschul et al., 2005). The putative binding sites for transcriptional factors nearby the core promoter sequences of *ldh1*, *ldh2*, and *ldh3* were identified with Yeastract (Teixeira et al., 2018).

2.4. Detection and expression of the genes ldh

2.4.1. Detection of the genes ldh by PCR

Total genomic DNA, extracted and quantified as reported above, was diluted to the concentration of 20 ng/μL. To detect the putative genes encoding for LDH among the isolates, the primers listed in Table 2 were designed, exploiting the nucleotidic polymorphism present in the gene sequences.

Table 2
Primers used for amplification of the genes *ldh1* and *ldh2*.

Primer	Sequence (5'-3')	Amplicon size (bp)	T_m (°C)
LDH1-F	ATCCGGTCGGAAGTACCAAC	195	59.4
LDH1-R	TACTGGTGACTGACAGCGC		58.8
LDH2-F	TTCAAGTACCAGCATTCGG	258	54.5
LDH2-R	CGGGTCAGCTAAACAGTA		53.7

The amplification reactions were performed in a final volume of 20 μL containing: 1× DreamTaq Green Buffer (Thermo Fisher Scientific), 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 1 μM of each primer, 0.025 U/μL DreamTaq DNA Polymerase (Thermo Fisher Scientific), and 50 ng of genomic DNA. The amplification program comprised an initial incubation at 94 °C for 4 min, 30 cycles of denaturation at 94 °C for 1 min, annealing at the respective melting temperature (T_m) of each pair of primers reported in Table 2 for 30 s and extension at 72 °C for 1 min, and concluded with a final extension of 8 min at 72 °C. The amplifications were conducted in a Thermal Cycler 2720 (Applied Biosystems, Thermo Fisher Scientific). The PCR products were verified by electrophoresis on 1.0 % agarose gel stained with EuroSafe (Euroclone S.p.A., Pero, Italy) and using the marker O'GeneRuler DNA (Thermo Fisher Scientific).

2.4.2. RNA extraction and cDNA synthesis

Samples for total RNA extraction were taken in triplicate at three different times during the microvinification trial described in paragraph 2.2: T0 (inoculation; 0 h), T1 (half fermentation; 144 h) and T2 (end of fermentation; 288 h). RNA was purified according to Combina et al. (2012).

All samples were then diluted on DEPC-treated water to reach a concentration of 100 ng/μL and submitted to DNase treatment. This purification was carried out with the Turbo DNA-free kit (Thermo Fisher Scientific), following the manufacturer's instructions. The purified RNA was kept at -80 °C, until the synthesis of total cDNA was performed using the ImProm-II™ Reverse Transcriptase kit (Promega), according to the manufacturer's protocol. Nucleic acid yield, purity and integrity were checked with NanoDrop ND1000 UV–vis Spectrophotometer and by electrophoresis on a 1,5 % (w/v) agarose gel.

2.4.3. Gene expression analysis by quantitative real-time PCR

Plasmid DNA standards were constructed by cloning the PCR products of *L. thermotolerans* CBS 6340^T in the pGEMT-easy vector and by transformation of the high efficiency JM109 competent cells (Promega), according to the manufacturer's guidelines.

Isolation of plasmid DNA from transformants was conducted with GenElute™ Plasmid Miniprep kit (Sigma-Aldrich). The identity of the inserts was confirmed by PCR assays. Quantification of the standards was performed with the Qubit 4 Fluorometer. The number of plasmid copies per microliter was obtained from the concentration values using the molecular weight of the recombinant plasmid DNA according to Lee et al. (2006).

Quantitative real-time PCR (qPCR) was carried out with the cDNA obtained as described above as target and ten-fold serial dilutions of recombinant plasmid (from 3 to 3×10^8 copies) to generate the standard curve. The absolute quantity of target in the samples was estimated interpolating the C_t value obtained in qPCR with the standard curve.

Amplifications of the genes *ldh1* and *ldh2* were carried out in 25 μL reaction volume, containing 1 × SYBR Green Master Mix (Life Technologies), 0.4 μM of each primer, and 6 ng of cDNA or a suitable amount of plasmid. Amplification program included an initial

denaturation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 1 min, annealing at the respective melting temperature (T_m) of each pair of primers reported in Table 2 for 30 s and extension at 72 °C for 40 s, and concluded with a final denaturation of 10 s at 95 °C. The melting curve was conducted by raising the temperature from 60 to 95 °C with a gradient of 0.1 °C/s. The reactions were performed with technical replicates on Light Cycler® Nano (Roche, Pleasanton, USA).

2.4.4. Statistical analysis

Data from LA determination and qPCR were compared by one way ANOVA (ANalysis Of VAriance) and then subjected to post-hoc Tukey's HSD (Honestly Significant difference) test, using the software PAST (Hammer et al., 2001). The difference was considered significant with a p-value below 0.05.

3. Results and discussion

3.1. Intraspecific variability in L(+)-lactic acid production

Eleven strains of *L. thermotolerans*, originated from diverse wine grape varieties, the type strain CBS 6340^T, and a commercial starter of the same species (CONCERTO) were evaluated for LA production, a distinctive feature of this yeast. The analysis was conducted by growing the strains in standardized conditions, i.e. using mYDP laboratory medium containing 220 g/L glucose. The growth curves (Fig. 1A), determined by measuring the OD_{600}, show that all strains had a similar behaviour in the liquid medium, and were reaching the early stationary growth phase after 72 h. On the other hand, a remarkable variability among the strains was observed in the LA production, measured after 72 h of growth, as shown in Fig. 1B.

LA production ranged from 1.15 g/L (COLC27) to 5.24 g/L (SOL13). The strains were divided into three groups according to the amount of LA produced, namely: low producers (COLC27, CBS 6340^T, CONCERTO, COLC11 and DESP53), medium producers (GLERA15, MALV17, LT3, FIANO43, LS15 and MALV13), and high producers (LT15 and SOL13). In our previous research (Binati et al., 2019), the same strains grown in Trebbiano grape must (210 g/L sugar) showed a variation in LA production from 0.26 g/L (CONCERTO) to 10.54 g/L (SOL13), with the same trend observed for the differential production among the strains. Three of them, further investigated in sequential fermentations with *S. cerevisiae* EC 1118 in Pinot Grigio grape must (236 g/L sugar), still retained their behaviour regarding LA production: 0.53 g/L for COLC27, 1.22 g/L for DESP53, and 4.42 g/L for SOL13 (Binati et al., 2020).

Recent researches have highlighted the intraspecific heterogeneity of *L. thermotolerans* in the LA accumulation. Hranilovic et al. (2018) tested 94 strains collected from diverse ecological niches worldwide and the LA production ranged from 1.8–12.0 g/L. Other previous studies with diverse strains found concentration values between 0.90–4.20 g/L in Viura pasteurized grape must (206 g/L sugars) (Escribano et al., 2018), 1.24–2.56 g/L in Aglianico grape must (240 g/L sugars) (Aponte and Blaiotta, 2016) and 1.0–16.6 g/L in Greek pasteurized grape must (215 g/L sugars) (Banilas et al., 2016).

Despite this well-documented variability, scarce information is available on the molecular mechanisms behind these metabolic pathways in *L. thermotolerans*. Aiming to further analyse the possible reasons for the huge phenotypic divergence on the LA production, a genome comparison was performed considering the lowest (COLC27) and the highest (SOL13) LA producers. It was focused on a molecular characterization of the lactate dehydrogenases and permeases involved in LA metabolism.

3.2. General overview of the L. thermotolerans COLC27 and SOL13 genome

3.2.1. Genome assembly

The genomes of the *L. thermotolerans* strains COLC27 and SOL13 were sequenced using Illumina NextSeq 500 platform at about 80X (STD 209) and 117X (STD 308) coverage. To date, only the complete genome of *L. thermotolerans* CBS 6430^T has been determined for this yeast species (Génolevures Consortium et al., 2009).

The general genomic features of COLC27 and SOL13 are summarized in Table 3.

Final assemblies had total sizes in the range of 10.2 and 10.3 Mb, consisting of 43 and 71 scaffolds for *L. thermotolerans* COLC27 and SOL13, respectively. The GC content of 47.34 % was the same previously identified for the type strain (47.3 %) (Génolevures Consortium et al., 2009).

The high quality of the assemblies was evidenced by the high N50 and N75 values (Table 3) and allowed a reliable gene finding and functional annotation. Further, the genome completeness of the strains SOL13 and COLC27, 99.3 % and 99.0 % respectively, was similar to the type strain (99.0 %), according to the presence of single copy universal fungi orthologs genes from BUSCO database.

In the COLC27 and SOL13 genomes there are, respectively, 6,227 and 6,356 protein-encoding genes, including 7,876 and 7,974 exons. As for other unicellular fungal genomes, the number of exons per gene is close to one, with a small number of genes predicted to have one or more introns. Comparative analysis performed between the two strains sequenced in the present study revealed the presence of 91,370 variants, 87,968 of them are SNPs (single-nucleotide polymorphisms), 3,402 are InDels (Insertion–deletion mutations). A bioinformatics prediction of the functional effect of these variants performed using SNPeff

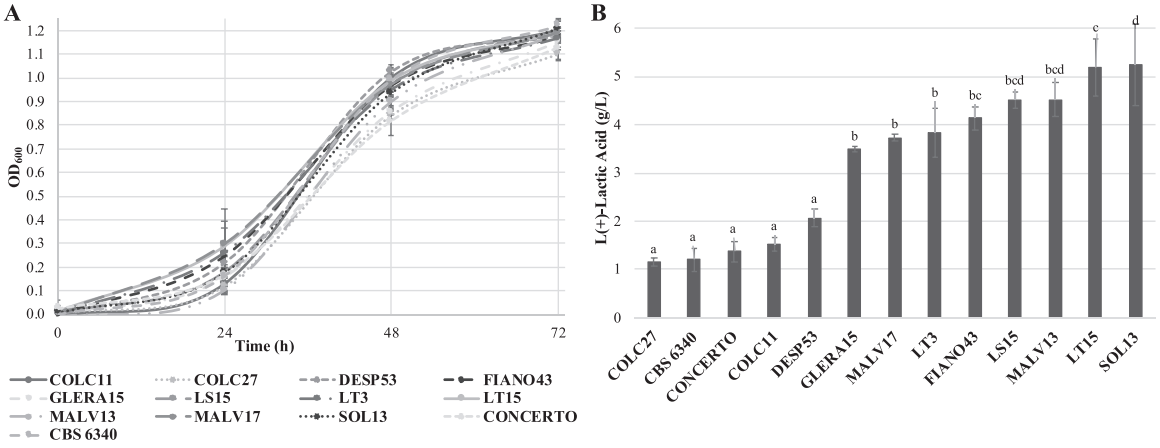

Fig. 1. Variability of the *L. thermotolerans* strains grown in mYPD medium. (A) Growth curves of the strains obtained by OD_{600}. (B) L(+)-lactic acid production after 72 h; values with different letters are significantly different ($p < 0.05$).

Table 3
Whole-genome information of the *L. thermotolerans* strains COLC27 and SOL13.

		L. thermotolerans COLC27	*L. thermotolerans* SOL13
Nucleotide distribution	GC (%)	47.34	47.34
Contig measurements	N75	478,272 bp	364,893 bp
	N50	788,172 bp	946,139 bp
	Scaffolds	43	71
Total length		10.2 Mb	10.3 Mb
Genome completeness		99.0 %	99.3 %
Protein coding genes		6,227	6,356
Exons		7,876	7,974

revealed that only 156 have a putative high phenotypic impact, while other 49,331 have moderate or low predicted effect.

Fig. 2A shows the annotated genes of the two *L. thermotolerans* strains grouped into functional categories, according to their biological roles. Annotation of the reference genome of CBS 6340^T is also reported for comparison. The annotation showed very similar results for the three strains, which have 55–58 complete KEGG modules, 15–17 of them associated to Carbohydrate metabolism (including for example the glyoxylate cycle). The Genetic Information Processing category accounted for the majority of KEGG annotation for all genomes. Other pathways including a high number of genes were Cellular Processes (241 − 246 genes) and Carbohydrate metabolism (195–199 genes). Moreover, most of KEGG-annotated genes were shared by the three strains, for a total of 2,823, much higher than the specific genes of each strain (Fig. 2B). Interestingly, a higher number of common genes was observed between COLC27 and CBS 6340^T than with SOL13.

The high similarity in terms of gene content among the strains suggested that the phenotypic differences were probably determined by single nucleotide or structural variants present in the genome and not to presence/absence of specific genes. Analysis of the genes where high impact variants have been identified revealed that many of them were associated to carbohydrate (K09101) and amino acids metabolism (K09105). In particular, the average number of variants for gene is 97, but some genes have more than 200 different variants. The variants annotation revealed that the latter are specifically linked with lactate metabolism (K00016) and arginine biosynthesis (K00220).

The alignment between the *L. thermotolerans* COLC27 and SOL13 genomes generated with MAUVE software did not reveal any translocations.

The availability of whole genomes allowed the reconstruction of a phylogenomic tree. A phylogenomic Maximum-likelihood tree (Fig. 3) was built based on 154 concatenated genes arranged in 11 orthologous clusters. The tree topology confirmed the 26*S* rRNA taxonomic assignment of the *L. thermotolerans* strains COLC27 and SOL13. Two distinct clades were recognized for the *Lachancea* species, where *L. dasiensis*, *L. nothofagi*, *L. meyersii*, *L. lanzarotensis* are gathered in the first clade. *L. thermotolerans* belongs to the second clade, together with *L. mirantina*, *L. kluyveri*, *L. fermentati*, *L. waltii* and *L. quebecensis*. The latter established a sister group of *L. thermotolerans*, revealing a proximal relationship of these yeasts.

3.3. Characterization of loci related to L(+)-lactic acid metabolism

Firstly, the genes involved in LA metabolism among the *L. thermotolerans* strains COLC27, SOL13, and CBS 6340^T were assessed, in order to establish if the differences can help to understand the variability in LA production.

Three different copies of the gene encoding for the LDH enzyme were identified from *in silico* analysis of the published genome sequence of CBS 6340^T, denominated here as *ldh1* (KLTH0D00440 g, Chromosome D), *ldh2* (KLTH0G19536 g, Chromosome G), and *ldh3* (KLTH0G19558 g, Chromosome G). It was previously demonstrated that LA production efficiency is affected not only by the enzymatic properties of LDH originated from different organisms, but also by the copy number of the gene *ldh* (Ilmén et al., 2013). Nevertheless, three copies of the gene *ldh* were also identified in the genomes of the two new *L. thermotolerans* strains studied. Further, as in the reference genome, the strains displayed the same *ldh* organization with two copies of *ldh* in tandem (*ldh2* and *ldh3*) and one (*ldh1*) located elsewhere.

To investigate the presence of the two known lactate permease genes *jen1* and *ady2*, the Pfam predicted proteins (PF00083, PF01184) were targeted in the GenBank file of the *L. thermotolerans* annotated genomes. A unique sequence of the *jen1* was found in the strains SOL13 and COLC27, while two copies of the permease *ady2* were found. On the other hand, the type strain owns one copy of each gene.

Alignment-based neighbour-*joining* method (Fig. 4) showed a higher correlation between the amino acid sequences of the LDHs and JEN1 derived from the low LA producer strains (COLC27 and CBS 6340^T), than those derived from the high producer (SOL13). These findings agree with the phenotypic results previously observed, where SOL13 produced, in mYPD broth, 4.5 times more LA than COLC27 and CBS 6340^T (Fig. 1B).

Nevertheless, it seems that differences in the amino acid sequences do not interfere with key sites in the proteins, which remained unchanged. This was verified by the alignment and thoroughly comparison of translated sequences of all genes taken in consideration. Some putative conserved domains were observed in the genes *ldh* (Supplementary Figure S.1): i. **GXGXXG** (Madern, 2002), ii. NAD(P) binding site, iii. substrate binding site, iv. dimer and tetramer interface.

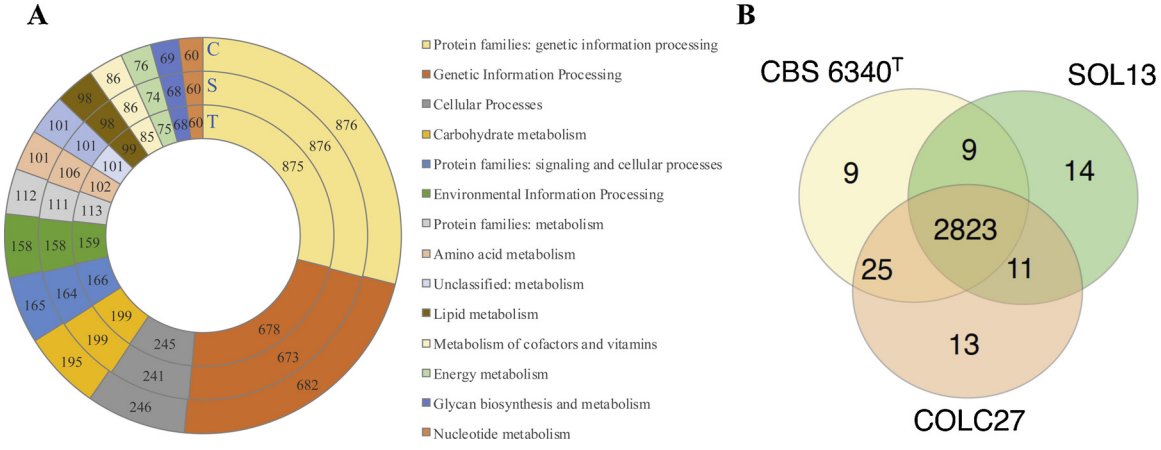

Fig. 2. (A) Functional annotation of the genome sequence of the three analysed *L. thermotolerans* strains obtained with the BlastKOALA tool; from the outer to the inner: COLC27; SOL13 and CBS 6340^T. (B) Venn diagram representing the common and specific annotated genes among the three strains.

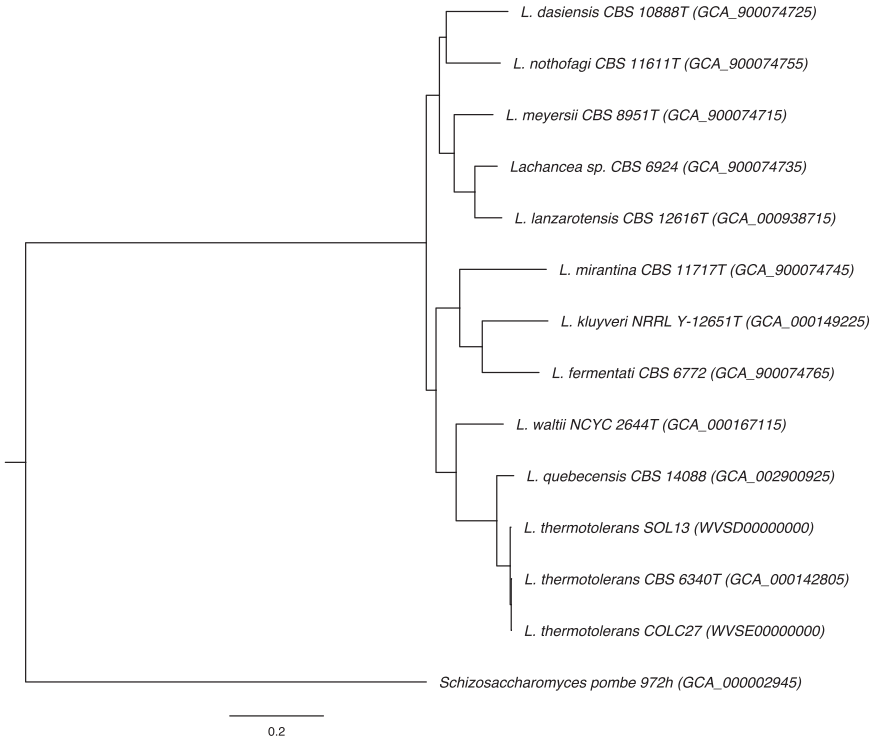

Fig. 3. Phylogenomic tree of the genus *Lachancea*, containing all genomes available in public databases and the two new genomes of the *L. thermotolerans* strains COLC27 and SOL13. The strain *S. pombe* 972 h was used as outgroup.

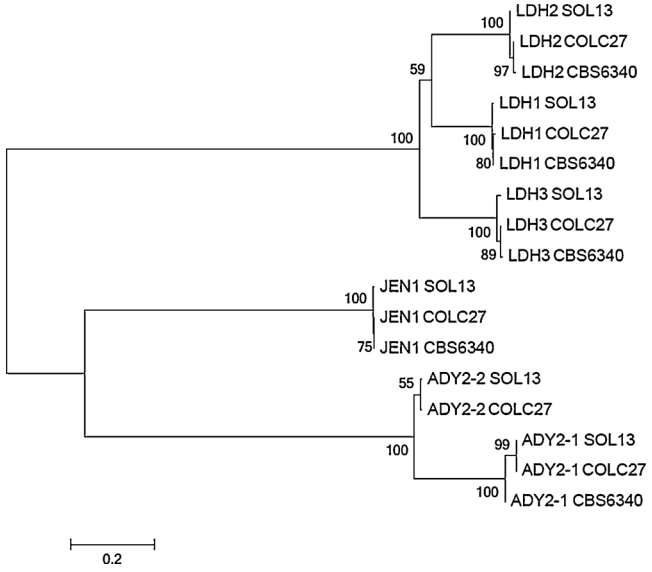

Fig. 4. Neighbour-joining tree of *L. thermotolerans* main known genes involved in L(+)-lactic acid metabolism.

Meanwhile, the sequence ^{379}NXX[S/T]HX[S/T]QDXXXT^{391}, part of the substrate translocation pathway (Soares-Silva et al., 2007), was found in *jen1* (Supplementary Figure S.2).

As regard *ady2*, the single copy shared by all three strains had a higher similarity between SOL13 and COLC27. It cannot be inferred whether the additional copy of *ady2*, present in the two selected strains, but not identified in the type strain, could provide any evolutionary advantage or impact on the LA metabolism. Considering the phenotypic results of LA production, it is most likely not the case, since COLC27 and CBS 6340^{T} had similar behaviour, and SOL13 was significantly different. Indeed, previous studies reported that overexpression of both *jen1* and *ady2* in an engineered strain of *S. cerevisiae* did not result in higher LA yields from glucose (Turner et al., 2019).

The presence of *jen*1 and *ady2* in all three genomes indicates a fully functional system of LA export. Moreover, considering that over-expression of such genes has been suggested in literature to provide only limited benefits in LA yields when using glucose as carbon source (Turner et al., 2019), it was decided to further funnel the analysis on the transcription of the genes *ldh*.

Regulatory sequences in the promoter region determines the expression pattern of a given gene, which directly influences the encoded protein activity level, therefore the upstream sequences of the genes *ldh* were compared and scrutinised. The analysis considered around 400 bases before the start codon (ATG), in order to incorporate the core promoter and upstream regulatory elements (Erb and van Nimwegen, 2011; Lubliner et al., 2015). The phylogenetic analysis of the promoter sequences (Fig. 5) showed a higher similarity between the *ldh1* and *ldh2* upstream regions. Interestingly, the promoters of *ldh1* were very similar among the three strains, while for *ldh2* and *ldh3*, a clear separation was

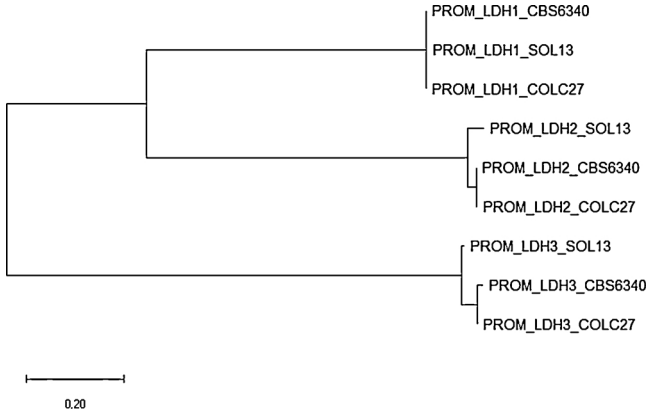

Fig. 5. Neighbour-joining tree of *L. thermotolerans* promoter regions upstream of the genes *ldh*.

witnessed for SOL13. Accordingly, the differences observed for the *ldh* promoters followed the trend stated for the *ldh* coding sequences (Fig. 4).

Going into more detail, the alignment of nucleotides in the *ldh* promoter sequences of the three investigated strains allowed the underlining of a conserved region located at around 100 base pairs upstream of the ATG. It was identified as the TATA box (Supplementary Figure S.3), a well-known core promoter element that, in yeasts, characterizes genes with variable expression and often induced in response to various stresses (Erb and van Nimwegen, 2011). Moreover, some single nucleotide mutations in the sequence of SOL13 were highlighted, since, as discussed by Lubliner et al. (2015), sequence variations within the core promoter regions greatly affect promoter activity.

The upstream regulatory elements typically contain one or more transcription factor binding sites (TFBSs). These short sequences (6–12 bp) play a part in regulating the expression of the associated genes through the recruitment of transcription factors (TFs) to the core promoter (Levo and Segal, 2014; Lubliner et al., 2015). The TFs interact with one another and with the transcriptional machinery, acting as repressors or activators of transcription, thus being very effective ways to alter a gene's expression and to balance biosynthetic pathways (Decoene et al., 2019).

Analysis of TFBSs for each *ldh* intergenic region considered in the COLC27, SOL13 and CBS 6340^T genomes (Supplementary Figure S.4) revealed very similar distribution profiles in the *ldh1* and *ldh3* promoters. On the other hand, the profile of TFBSs in *ldh2* is different across the compared strains, which could be related with mutations during their evolution. TFBSs involved in transcription regulation are usually distributed in the first 200 base pairs upstream of transcriptional start sequence (Erb and van Nimwegen, 2011). In this precise *ldh2* promoter region, four TFBSs were detected exclusively in SOL13 (Pip2, Snk7, Ime1, Hsf1) while another one (Rtg1/Rtg3) was present only in COLC27 and CBS 6340^T, despite the large presence of TFBSs across the sequences. TFs are usually associated with regulation mechanisms involved in environmental stress response (osmolarity, acidity, heat, starvation), but, to the best of our knowledge, no information are available on TFs related to regulation of genes in LA metabolism. Nevertheless, the observed differences in TFBSs profiles, and the single nucleotide mutations in the core promoters of SOL13, suggest that the divergent strain behaviour could be determined by an altered regulation of the genes *ldh*.

Thus, to check whether the observed similarities and differences in promoter analysis could result in significant transcription changes, the genomic comparison was refined to the level of gene expression for *ldh1* and *ldh2*.

3.4. Microvinification trials and expression analysis

The qPCR assay was used to monitor the expression of the genes *ldh1* and *ldh2* through a microvinification trial of pasteurized grape must, inoculated with the strains COLC27 and SOL13. Samples for RNA extraction were collected at the beginning (T0, 0 h), middle (T1, 144 h) and at the end (T2, 288 h) of fermentation, as determined by weight loss method. Both strains grew well and showed a similar growth curve (data not shown). The strains fermentation kinetics (Fig. 6A) showed different fermentative ability. SOL13 started the sugar consumption promptly, while COLC27 was slower in the first half of fermentation. After the sixth day, COLC27 increased its fermentation rate and both strains were able to conclude the process 12 days after the inoculation.

As it can be seen in Fig. 6B, the high-producer SOL13 reached approximately 11 times more LA than the low-producer COLC27. Noteworthy, the LA levels reached the maximum value in the middle of fermentation (T1), since the concentration did not augment significantly at the end (T2) of the process. LA production during the first half of fermentation was also observed by other authors following this metabolic activity during the course of fermentation (Morata et al., 2019). The direct conversion of intracellular pyruvate to LA, instead of producing ethanol, could represent an advantageous alternative pathway for NAD^+ regeneration. Nonetheless, extracellular accumulation of the produced lactate and the consequent reduction of pH create a threshold limiting LA production (Sauer et al., 2010). Moreover, this threshold varied among the analyzed strains.

The transcription levels of the genes *ldh1* and *ldh2* for the two strains are shown in Fig. 7.

In general, the expression level differed considerably depending on the gene, the strain, and the fermentation course, as well. The values obtained at T0 for both genes were not reported in the graph, because they reflected the residual expression associated with the stationary phase in the YPD pre-culture conditions. Regarding the gene *ldh1*, in both strains its transcription occurred to a greater extent in the middle of fermentation (T1), suggesting a higher activity during the phase of tumultuous fermentation and then it declined towards the end of the process. The strain SOL13 displayed a significantly higher expression of the gene *ldh1* at T1 than COLC27, up to 7.10 log and 6.39 log (copies/μg cDNA), respectively.

A similar trend was shown for *ldh2*, with an even higher difference of 2 log between the transcript levels for the two strains in the mid-fermentation, 6.36 log(copies/μg cDNA) for SOL13 and 4.34 log(copies/μg cDNA) for COLC27. The expression significantly dropped at the end of the process, markedly for SOL13.

The results of qPCR analysis are in accordance with the phenotypic outcome shown in Fig. 6B, as LA production by SOL13 was much higher than COLC27 and this metabolic pathway was active in the middle of the fermentation process. After that, the concentration of LA did not change, in agreement with the much lower level of transcripts found at this phase. Moreover, they confirm the previous suggestion of a possible correlation of this metabolism with the transcriptional regulation of the genes *ldh*.

A recent investigation of five vineyard-associated *L. thermotolerans* strains (plus the type strain), with varying LA production in filter-sterilized grape must, quantified the relative expression of the three genes *ldh* after 16 h. It was highlighted a significantly higher expression of *ldh*2 in the top-producers, in agreement with the present results. Additionally, they did not find differences in *ldh*1 and *ldh*3 expression, suggesting that regulation of only *ldh*2 is related to LA production (Sgouros et al., 2020).

The improved LA production level could be in principle associated with a better transcription rate of the mutated genes *ldh* in SOL13. Notwithstanding, other outcomes of the sequence polymorphisms should be considered, such as a lower turnover rate or a higher translation level of the transcript (Branduardi et al., 2006).

4. Conclusions

Even if a relatively low number of *L. thermotolerans* strains was studied, a remarkable phenotypic diversity was found for a technologically relevant pathway as LA production. Bioinformatics tools allowed genome annotation of an important yeast species, enriching the public databases with two strains greatly divergent in that character. The acquired genome information guided the analysis of key genes involved in the LA metabolism and their promoter regions. This approach allowed the design of an additional step of gene expression analysis, which gave further evidence of the regulatory mechanisms most likely related to the variability investigated.

Looking forward, new studies are required to broaden the understanding on this matter. Metabolomics and transcriptomics could be very powerful tools when coupled with genome-wide analysis to uncover additional metabolic pathways and genes involved in carbon flux, redox balance, and stress responses triggered by LA accumulation. Aspects of cellular growth and interactions with other microorganisms could also be relevant to better exploit *L. thermotolerans* intraspecific

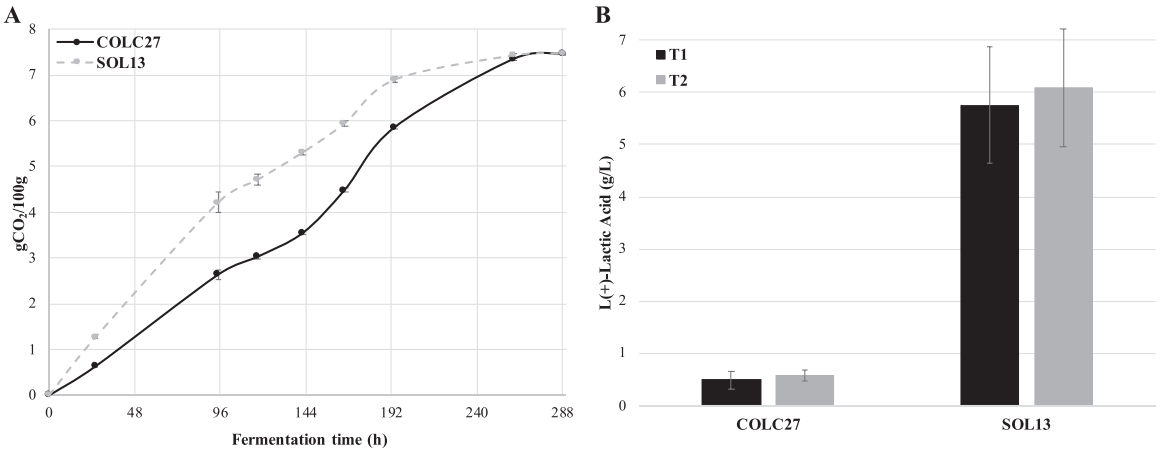

Fig. 6. Variability of the *L. thermotolerans* strains COLC27 and SOL13 in pasteurized grape must. (A) Fermentation kinetics. (B) Amount of L(+)-lactic acid produced in the middle (T1, 144 h) and at the end (T2, 288 h) of fermentation.

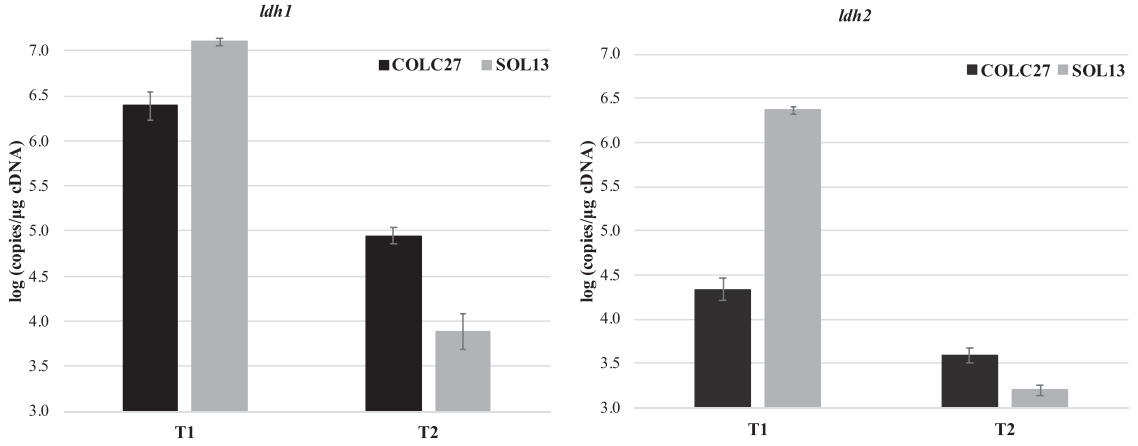

Fig. 7. Expression of the genes *ldh1* and *ldh2* for the *L. thermotolerans* strains COLC27 and SOL13 in the middle (T1, 144 h) and at the end (T2, 288 h) of fermentation in pasteurized grape must.

variability for different biotechnological applications.

CRediT authorship contribution statement

Veronica Gatto: Investigation, Formal analysis, Writing - original draft. **Renato L. Binati:** Investigation, Formal analysis, Writing - original draft. **Wilson J.F. Lemos Junior :** Conceptualization, Formal analysis. **Arianna Basile:** Formal analysis. **Laura Treu:** Formal analysis. **Otávio G.G. de Almeida:** Formal analysis. **Giada Innocente:** Investigation. **Stefano Campanaro:** Resources, Supervision. **Sandra Torriani:** Conceptualization, Resources, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Francesca Cacciatori for the assistance in laboratory trials. The PhD scholarship of R.L.B. was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), grant number GDE 202265/2015-3.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.micres.2020.126525.

References

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schäffer, A.A., Yu, Y.-K., 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 272, 5101–5109. https://doi.org/10.1111/j.1742-4658.2005.04945.x.

Aponte, M., Blaiotta, G., 2016. Potential role of yeast strains isolated from grapes in the production of Taurasi DOCG. Front. Microbiol. 7, 809. https://doi.org/10.3389/fmicb.2016.00809.

Axelsson, L., 2004. Lactic acid bacteria: classification and physiology. In: Salminen, S., Von Wright, A., Ouwehand, A. (Eds.), Advance in Lactic Acid Bacteria: Microbiological and Functional Aspects. Marcel Dekker Inc., New York, pp. 1–66. https://doi.org/10.1201/9780824752033.

Balikci, E.K., Tanguler, H., Jolly, N.P., Erten, H., 2016. Influence of *Lachancea thermotolerans* on cv. Emir wine fermentation. Yeast 33, 313–321. https://doi.org/10.1002/yea.3166.

Banilas, G., Sgouros, G., Nisiotou, A., 2016. Development of microsatellite markers for *Lachancea thermotolerans* typing and population structure of wine-associated isolates. Microbiol. Res. 193, 1–10. https://doi.org/10.1016/j.micres.2016.08.010.

Benito, S., 2018. The impacts of *Lachancea thermotolerans* yeast strains on winemaking. Appl. Microbiol. Biotechnol. 102, 6775–6790. https://doi.org/10.1007/s00253-018-9117-z.

Benito, A., Calderon, F., Palomero, F., Benito, S., 2016. Quality and composition of Airén wines fermented by sequential inoculation of *Lachancea thermotolerans* and *Saccharomyces cerevisiae*. Food Technol. Biotechnol. 54, 135–144. https://doi.org/10.17113/ftb.54.02.16.4220.

Besemer, J., Borodovsky, M., 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33, 451–454. https://doi.org/10.1093/nar/gki487.

Binati, R.L., Innocente, G., Gatto, V., Celebrin, A., Polo, M., Felis, G.E., Torriani, S., 2019. Exploring the diversity of a collection of native non-*Saccharomyces* yeasts to develop co-starter cultures for winemaking. Food Res. Int. 122, 432–442. https://doi.org/10.

1016/j.foodres.2019.04.043.
Binati, R.L., Lemos Junior, W.J.F., Luzzini, G., Slaghenaufi, D., Ugliano, M., Torriani, S., 2020. Contribution of non-*Saccharomyces* yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia* spp. and *Starmerella bacillaris* strains isolated in Italy. Int. J. Food Microbiol. 318, 108470. https://doi.org/10.1016/j.ijfoodmicro.2019.108470.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
Branduardi, P., Sauer, M., De Gioia, L., Zampella, G., Valli, M., Mattanovich, D., Porro, D., 2006. Lactate production yield from engineered yeasts is dependent from the host background, the lactate dehydrogenase source and the lactate export. Microb. Cell Fact. 5, 4. https://doi.org/10.1186/1475-2859-5-4.
Capozzi, V., Garofalo, C., Chiriatti, M.A., Grieco, F., Spano, G., 2015. Microbial terroir and food innovation: the case of yeast biodiversity in wine. Microbiol. Res. (Pavia) 181, 75–83. https://doi.org/10.1016/j.micres.2015.10.005.
Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B.A., Wang, Z., 2013. Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production. Biotechnol. Adv. 31, 1200–1223. https://doi.org/10.1016/j.biotechadv.2013.02.009.
Cingolani, P., Platts, A., Wang, l., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6, 80–92. https://doi.org/10.4161/fly.19695.
Combina, M., Pérez-Torrado, R., Tronchoni, J., Belloch, C., Querol, A., 2012. Genome-wide gene expression of a natural hybrid between *Saccharomyces cerevisiae* and *S. kudriavzevii* under enological conditions. Int. J. Food Microbiol. 157, 340–345. https://doi.org/10.1016/j.ijfoodmicro.2012.06.001.
Contreras-Moreira, B., Vinuesa, P., 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol. 79, 7696–7701. https://doi.org/10.1128/AEM.02411-13.
Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403. https://doi.org/10.1101/gr.2289704.
Datta, R., Henry, M., 2006. Lactic acid: recent advances in products, processes and technologies - a review. J. Chem. Technol. Biotechnol. 81, 1119–1129. https://doi.org/10.1002/jctb.1486.
de Kok, S., Nijkamp, J.F., Oud, B., Roque, F.C., de Ridder, D., Daran, J.-M., Pronk, J.T., van Maris, A.J.A., 2012. Laboratory evolution of new lactate transporter genes in a jen1Δ mutant of *Saccharomyces cerevisiae* and their identification as ADY2 alleles by whole-genome resequencing and transcriptome analysis. FEMS Yeast Res. 12, 359–374. https://doi.org/10.1111/j.1567-1364.2011.00787.x.
Decoene, T., De Maeseneire, S.L., De Mey, M., 2019. Modulating transcription through development of semi-synthetic yeast core promoters. PLoS One 14, e0224476. https://doi.org/10.1371/journal.pone.0224476.
Domizio, P., House, J.F., Joseph, C.M.L., Bisson, L.F., Bamforth, C.W., 2016. *Lachancea thermotolerans* as an alternative yeast for the production of beer. J. Inst. Brew. 122, 599–604. https://doi.org/10.1002/jib.362.
Erb, I., van Nimwegen, E., 2011. Transcription factor binding site positioning in yeast: proximal promoter motifs characterize TATA-less promoters. PLoS One 6, e24279. https://doi.org/10.1371/journal.pone.0024279.
Escribano, R., González-Arenzana, L., Portu, J., Garijo, P., López-Alfaro, I., López, R., Santamaría, P., Gutiérrez, A.R., 2018. Aromatic compound production and fermentative behavior within different non-*Saccharomyces* species and clones. J. Appl. Microbiol. 124, 1521–1531. https://doi.org/10.1111/jam.13735.
Freel, K.C., Friedrich, A., Hou, J., Schacherer, J., 2014. Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. Genome Biol. Evol. 6, 2586–2594. https://doi.org/10.1093/gbe/evu203.
Génolevures Consortium, Souciet, J.L., Dujon, B., Gaillardin, C., Johnston, M., Baret, P.V., Cliften, P., Sherman, D.J., Weissenbach, J., Westhof, E., Wincker, P., Jubin, C., Poulain, J., Barbe, V., Ségurens, B., Artiguenave, F., Anthouard, V., Vacherie, B., Val, M.E., Fulton, R.S., et al., 2009. Comparative genomics of protoploid *Saccharomycetaceae*. Genome Res. 19, 1696–1709. https://doi.org/10.1101/gr.091546.109.
Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electronica 4, 1–9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
Hranilovic, A., Bely, M., Masneuf-Pomarede, I., Jiranek, V., Albertin, W., 2017. The evolution of *Lachancea thermotolerans* is driven by geographical determination, anthropisation and flux between different ecosystems. PLoS One 12, e0184652. https://doi.org/10.1371/journal.pone.0184652.
Hranilovic, A., Gambetta, J.M., Schmidtke, L., Boss, P.K., Grbin, P.R., Masneuf-Pomarede, I., Bely, M., Albertin, W., Jiranek, V., 2018. Oenological traits of *Lachancea thermotolerans* show signs of domestication and allopatric differentiation. Sci. Rep. 8, 14812. https://doi.org/10.1038/s41598-018-33105-7.
Ilmén, M., Koivuranta, K., Ruohonen, L., Rajgarhia, V., Suominen, P., Penttilä, M., 2013. Production of L-lactic acid by the yeast *Candida sonorensis* expressing heterologous bacterial and fungal lactate dehydrogenases. Microb. Cell Fact. 12, 53. https://doi.org/10.1186/1475-2859-12-53.
Jolly, N.P., Varela, C., Pretorius, I.S., 2014. Not your ordinary yeast: non-*Saccharomyces* yeasts in wine production uncovered. FEMS Yeast Res. 14, 215–237. https://doi.org/10.1111/1567-1364.12111.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
Kurtzman, C.P., 2003. Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res. 4, 233–245. https://doi.org/10.1016/S1567-1356(03)00175-2.
Lachance, M.-A., Kurtzman, C.P., 2011. *Lachancea* kurtzman, *Lachancea* kurtzman (2003) In: Kurtzman, C.P., Fell, J.W., Boekhout, T. (Eds.), The Yeasts: A Taxonomic Study. Elsevier, New York, pp. 511–519. https://doi.org/10.1016/B978-0-444-52149-1.00041-0.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.
Lee, C., Kim, J., Shin, S.G., Hwang, S., 2006. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. J. Biotechnol. 123, 273–280. https://doi.org/10.1016/j.jbiotec.2005.11.014.
Lemos Junior, W.J.F., Binati, R.L., Felis, G.E., Slaghenaufi, D., Ugliano, M., Torriani, S., 2020. Volatile organic compounds from *Starmerella bacillaris* to control gray mold on apples and modulate cider aroma profile. Food Microbiol. 89, 103446. https://doi.org/10.1016/j.fm.2020.103446.
Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259. https://doi.org/10.1093/nar/gkz239.
Levo, M., Segal, E., 2014. In pursuit of design principles of regulatory sequences. Nat. Rev. Genet. 15, 453–468. https://doi.org/10.1038/nrg3684.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., Segal, E., 2015. Core promoter sequence in yeast is a major determinant of expression level. Genome Res. 25, 1008–1017. https://doi.org/10.1101/gr.188193.114.
Madern, D., 2002. Molecular evolution within the L-Malate and L-Lactate dehydrogenase super-family. J. Mol. Evol. 54, 825–840. https://doi.org/10.1007/s00239-001-0088-8.
Makuc, J., Paiva, S., Schauen, M., Krämer, R., André, B., Casal, M., Leão, C., Boles, E., 2001. The putative monocarboxylate permeases of the yeast *Saccharomyces cerevisiae* do not transport monocarboxylic acids across the plasma membrane. Yeast 18, 1131–1143. https://doi.org/10.1002/yea.763.
Martinez, F.A.C., Balciunas, E.M., Salgado, J.M., González, J.M.D., Converti, A., Oliveira, R.P.S., 2013. Lactic acid properties, applications and production: a review. Trends Food Sci. Technol. 30, 70–83. https://doi.org/10.1016/j.tifs.2012.11.007.
Minh, B.Q., Nguyen, M.A., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024.
Morata, A., Bañuelos, M.A., Vaquero, C., Loira, I., Cuerda, R., Palomero, F., González, C., Suárez-Lepe, J.A., Wang, J., Han, S., Bi, Y., 2019. *Lachancea thermotolerans* as a tool to improve pH in red wines from warm regions. Eur. Food Res. Technol. 245, 885–894. https://doi.org/10.1007/s00217-019-03229-9.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Osburn, K., Amaral, J., Metcalf, S.R., Nickens, D.M., Rogers, C.M., Sausen, C., Caputo, R., Miller, J., Li, H., Tennessen, J.M., Bochman, M.L., 2018. Primary souring: a novel bacteria-free method for sour beer production. Food Microbiol. 70, 76–84. https://doi.org/10.1016/j.fm.2017.09.007.
Pacheco, A., Talaia, G., Sá-Pessoa, J., Bessa, D., Gonçalves, M.J., Moreira, R., Paiva, S., Casal, M., Queirós, O., 2012. Lactic acid production in *Saccharomyces cerevisiae* is modulated by expression of the monocarboxylate transporters Jen1 and Ady2. FEMS Yeast Res. 12, 375–381. https://doi.org/10.1111/j.1567-1364.2012.00790.x.
Sauer, M., Porro, D., Mattanovich, D., Branduardi, P., 2008. Microbial production of organic acids: expanding the markets. Trends Biotechnol. 26, 100–108. https://doi.org/10.1016/j.tibtech.2007.11.006.
Sauer, M., Porro, D., Mattanovich, D., Branduardi, P., 2010. 16 years research on lactic acid production with yeast-ready for the market? *Biotechnol.* Genet. Eng. Rev. 27, 229–256. https://doi.org/10.1080/02648725.2010.10648152.
Sgouros, G., Mallouchos, A., Filippousi, M.E., Banilas, G., Nisiotou, A., 2020. Molecular characterization and enological potential of a high lactic acid-producing *Lachancea thermotolerans* vineyard strain. Foods 9, E595. https://doi.org/10.3390/foods9050595.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
Soares-Silva, I., Paiva, S., Diallinas, G., Casal, M., 2007. The conserved sequence NXX[S/T]HX[S/T]QDXXXT of the lactate/pyruvate:H^+ symporter subfamily defines the function of the substrate translocation pathway. Mol. Membr. Biol. 24, 464–474. https://doi.org/10.1080/09687680701342669.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. https://

doi.org/10.1093/molbev/mst197.
Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41. https://doi.org/10.1186/1471-2105-4-41.
Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., Cavalheiro, M., Antunes, M., Lemos, A., Pedreira, T., Sá-Correia, I., 2018. YEASTRACT, an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. Nucl. Acids Res. 46, D348–D353. https://doi.org/10.1093/nar/gkx842.
Turner, T.L., Lane, S., Jayakody, L.N., Zhang, G., Kim, H., Cho, W., Jin, Y., 2019. Deletion of JEN1 and ADY2 reduces lactic acid yield from an engineered *Saccharomyces cerevisiae*, in xylose medium, expressing a heterologous lactate dehydrogenase. FEMS Yeast Res. 19https://doi.org/10.1093/femsyr/foz050. foz050.
Witte, V., Krohn, U., Emeis, C.C., 1989. Characterization of yeasts with high L [+]-lactic acid production: lactic acid specific soft-agar overlay (LASSO) and TAFE-patterns. J. Basic Microbiol. 29, 707–716 1 https://doi.org/0.1002/jobm.3620291014.